Network Modeling Analysis in Health Informatics and Bioinformatics (2020) 9:23
https://doi.org/10.1007/s13721-020-00228-9

ORIGINAL ARTICLE

A comparison and analysis of the Twitter discourse related to weight loss and fitness

Shabnam Shadroo[1] · Mohsen Yoosefi Nejad[2] · Ahmed Omar Bali[3] · Mehdi Hosseinzadeh[4] · Maryam sadat Delghandi[5]

Received: 29 July 2019 / Revised: 18 January 2020 / Accepted: 17 March 2020

Abstract
More than 30% of the world population is concerned with the problem of overweight. Social media can play a role in human health by offering them correct food patterns and increasing their awareness about different features of appropriate food and diet. Several researches have been carried out on context analysis of social network messages, but there is a paucity of literature on analysis of feelings in tweets and their different geographical locations. This study aims at understanding tweets stated on the amount of reception shown by people in the course of weight loss in a period of 1 month. This study uses cross-sectional and descriptive method to analyze over 2,684,858 of tweets quantitatively. It also compares the emotional aspects present in the tweets. Users, who are active in this domain, are classified into six classes. An investigation and comparison of the number of activities with relation to weight loss has been carried out by searching users' geographical information of social networks in different continents. English tweets have been chosen because of the generality of the English language. After reviewing the previous literature and the results of the analysis on these tweets, using the MALLET software, six classifications were considered for the tweets. The results show that there is a meaningful relation among the extracted parameters in the research.

Keywords Weight loss · Fitness · Users' classification · Geographical classification · Emotion analysis

1 Introduction

A considerable number of people use online social networks to learn about different issues, specially the issue of health and correct life style (May et al. 2016). It is possible to reach various useful results by investigating social networks and analyzing the messages sent in them. In the present study, we will be investigating Twitter social network. According to the statistics in 2016, out of the 1.3 billion accounts in Twitter, 320 million have been active. About 500 million tweets each day and about 6000 tweets each second are sent on average. Almost 65% of American companies have used Twitter for marketing and 77% of the users, welcome Twitter's offer (Brandwatch 2016). Therefore, Twitter can be a fast comfortable and effective tool for sending messages, and Twitter users can share a tweet with millions of other users around the world. Among the various very popular subjects, the ones related to weight loss have been investigated in this research. Waring (2016) in his report mentions that 1.3 of American women aged between 20 and 39 years, are overweight and most women have experienced post-pregnancy obesity at least once in their life. Also due to people's use of

✉ Mehdi Hosseinzadeh
Hosseinzadeh.m@Iums.ac.ir

Shabnam Shadroo
sh_shadroo@mshdiau.ac.ir

Mohsen Yoosefi Nejad
m_yoosefi@pnu.ac.ir

Ahmed Omar Bali
ahmed.bali@uhd.edu.iq

Maryam sadat Delghandi
maryam.delghandi@iaus.ac.ir

1 Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran

2 Department of Computer Engineering and Information Technology, Payame Noor University, Tehran, Iran

3 Diplomacy and Public Relations Department, University of Human Development, Sulaymaniyah, Iraq

4 Health Management and Economics Research Center, Iran University of Medical Sciences, Tehran, Iran

5 Department of Computer Engineering, Sabzevar Branch, Islamic Azad University, Sabzevar, Iran

Published online: 01 April 2020

fast foods and convenience food in fast food restaurants or takeaways, people's weight in the society is on the increase, as a result, more than 30% of people are suffered from overweight (Richard Dobbs 2014). Large number of people including latescent and pregnant women and others are in search of useful information about healthy and low-calorie food. Most of these people are very busy to visit a nutrition specialist. As a result, Twitter can be used as an appropriate tool for offering a healthy diet and increasing users' awareness and consequently controlling health in society. The key questions in this research are first, which groups on Twitter are active in the field of weight loss, second, what the attitude of present users is, and finally, the users in which geographical region are more active on Twitter. In the following literature review, the related research will be presented to identify the gap which formulates the goal of the present study, which investigates the tweets on weight loss through using cross-sectional and descriptive method.

2 Related work

Various researches have been conducted with relation to weight loss each following a particular aspect. Different studies have also been done on health provision centers and the issue of weight loss. Gabrielle et al. for their research embarked on gathering tweets containing weight loss phrases in a year's period (2012), they analyzed the mentioning of weight loss, and the number of tweets. Additionally, they compared the number of tweets in three time periods of before, during and after the holidays winter and summer, before during and after new year (Gabrielle and Turner-McGrievy 2015).

May et al. (2016) tried to find weight loss supportive committees on social networks. In this research, four Twitter user accounts were made in a period of 2–5 weeks which were planned to follow specialist health centers and the users' accounts of their members. The results indicated that less than 10% of personal specialist accounts followed the created accounts in the research, and the number of interactions and followings didn't depend on the weight conditions of accounts. They found out that women need more than 5 weeks to follow the accounts related to weight loss.

Lydecker (2016) did a research on Twitter about the rude remarks related to people's weight such as fat and lazy. The data in this research were gathered in the time period between 4 and 12 pm on May 31, 2013. The results indicated that 5.5% of the data from that day contained such remarks. The results also showed that 56.57% of the messages had negative connotation and 32.09% were neutral.

Martinus Evans (2016) conducted a study investigating weight loss in bloggers where 92% of the people studied were women and the average age for them was 35. The subjects were chosen from Twitter and for completing the process, some were also chosen from Facebook. The average weight loss since the beginning of the plan was 42.3 lb which shows that bloggers marked a success in weight-loss plan.

Waring (2016) studied the weight-loss programs in women who had recently bore babies. To complete the plan, people aged 18–45 were used from which 40% were overweight and 83% intended to lose weight in this study 81% of users showed a tendency to have a certain weight loss plan on Twitter.

Gruver carried out a number of pilot studies on mothers who wanted to prevent obesity for having healthy children. Posts were designed with new information showing the behavior of parents with healthy children they also held an interview conference. A program containing different health points such as blood pressure, heart rate, brain wave, and temperature was developed. Clinical personnel, 97% of women and 87% of doctors preferred not to attend interviews while encouraging their patients to do so. During a period of 8 weeks, they had watched the weekly video and regularly sent 4.4 posts related to the topic. This method is fully acceptable for low-income mothers with babies above obesity line. And they can be added to the group through social networks (Gruver 2016).

With the purpose of upgrading healthy lifestyle, Safran Naimark (2015) conducted a series of studies on the effect of applied programs, and the program was conducted so as to monitor the amount of physical activity and their diet. More than 99 people took part of whom 85 people were present till the end, 56 and 29 people were studied. The average age was 47.9 and BMI was 26.2. The results of these 14 weeks indicated that new web-based applied programs influence the index of people lifestyle. More elaborate results will be obtained if longer is spent on the study.

Chou et al. did a similar research in the US. Almost 69% of the elderly in America have access to the internet. Approximately 5% are members of support groups through internet, 7% in weblogs, and 23% use social networking websites (Chou 2009).

A study conducted in Link Medical Heart center in Denmark also highlighted the necessity of web-based applied programs, believing that people need the support from these programs to enhance healthy lifestyle (Blei 2003).

Gabrielle's analysis is only about the number of tweets and the mentioning of weight loss, but it lacks other issues such as geographical situation and feelings. The present study has tried to cover the analysis of more varied issues concerning weight loss.

Investigating the available literature reveals that, as far as we know, none of them has exclusively analyzed the results of feeling, content of tweets, and their geographical location.

The present research analyses the tweets in terms of contents, feelings, and users locations.

3 Method

Quantitative approach through a cross-sectional analysis method has been used in analyzing tweet data. Tweets related to weight loss were gathered during the time periods December 27, 2017, and January 30, 2018. Twitter API has been used which gathers tweets online. Using the key words #health, #diet, #fitness, #weightloss, obesity, weight lose attempt, weight lose journey. 2,684,858 tweets were gathered from 545,524 Twitter users. The hashtags used in the study were taken from hashtagify.me which was used in Gabrielle and Turner-McGrievy (2015).

The analysis presented in the study includes the following stages.

1. Pre-processing the data
2. Analyzing and grouping users
3. Sorting out the topics of the tweets
4. Analyzing tweets in a time period
5. Analyzing attitudes and feelings
6. Analyzing geographical location

3.1 Pre-processing the data

In the pre-processing phase, first a filter was placed on the tweets so that only those tweets including fitness and weight loss remained. After excluding the unrelated tweets, the users who had sent those tweets were also removed from the population. Then the tweets with insufficient information and the ones including non-English words were omitted. For this purpose, the percentage of English words compared to other words was calculated. The tweets with more than 60% of useful information were analyzed. The number of related tweets after omission totaled 1,673,559.

3.2 Analyzing and grouping users

After getting the data from Twitter API, the total number of users was 545,524 which were then reduced to 355,856, leaving out those users whose tweets had been omitted in the pre-process phase. Thus the subjects were narrowed down to only those who focused on weight loss and fitness.

The sent tweets 1,243,996 were retweets and 429,563 original ones.

In this phase, Twitter users were categorized into six groups. The groups include commercial centers, research centers, news centers, medical centers, celebrities, and common people. The users in whose profile the keywords from Table 2 have been used will join these centers. Also for commercial, research, medical, and news centers, the number of followers was more than 10,000 and for celebrities it was more than 100,000 common users are those who cannot be placed in these groups.

The words in Table 1 were searched in users' specification table in the fields of name, screen name, and description, so that users' groups were specified. Using these words, we can evaluate research centers which are active in the field of weight loss and fitness or medical centers which aim at introducing weight-loss medicine or commercial centers which are marketing and advertising weight loss and fitness products. Table 2 shows the number of members of each group.

As shown in the tables, the number of the group related to common people has been more than all other groups and next is the group related to group centers.

Table 2 The number of members of each group

Name of category	Number of users	Name of category	Number of users
Research center	4525	Business center	845
News center	4082	Medical center	365
Celebrity	767	Individual	348,272

Table 1 Keywords for grouping users

Name of category	Keywords
Research center	Foundation, association, organization, research, community, university, committee, community, founder, science, official, technology, tech, education, groupsociety, committee, volunteer, country, government, firm, company, companies, info, care thing
News Center	News, TV, channel, social, media, Facebook, radio, Instagram, information, magazine, network, show, video, site, Youtube, group, society, journalism, snapchat, web, social
Business center	Marketing, business, promotions, publicity, advertising, agency
Medical center	Clinic, hospital, medical, center, health care, fitness center, medicine, pharmacy, pharmaceutical
Celebrity	Celebrity, actor, actress, music, blogger, author, writer, influencer, artist, director, international manager, expert

3.3 Sorting out the topics of the tweets

In this section, the tweets were classified according to content. For this purpose, MALLET software (McCallum 2002) which is an open-source software and uses the LDA method for sorting out the topics was applied. The output of this software can help find the main topics of the text. First, the dataset which were in MySQL format was converted to text and non-standard characters were removed from text file. Then the text file is fed into mallet software and the LDA algorithm executes. The output was a list of topics which contains some related words but without any topic title. Based on the meaning of related words, a topic title was selected by researchers. For example, a topic which contained the words cancer, illness, diabetes, risk, and epidemic decided to be disease topic. Some topics contained similar words and were merged together by researchers. After adjusting topics, the related words of each topics were counted in every tweet. The tweet was labeled by a topic with maximum frequency of occurrence of related words. For example, if the tweet contains four words of topic 1 and eight words of topic 2, it was labeled as topic 2. The content is classified using the stated keywords in Table 3.

3.4 Analyzing the topics in a period of time

This phase looks into the number of tweets related to the topics created in the previous section and their number on a daily and weekly basis.

The data are collected during December 27 and January 30 and Tuesday is first day of data collection. For the purpose of weekly analysis of data, they are divided in to time periods of Tuesday to Tuesday during 5 weeks.

3.5 Analyzing attitude and feelings

Opinion mining is a recent subdiscipline at the crossroads of information retrieval and computational linguistics which is concerned not with the topic a document is about, but with the opinion it expresses (Esuli and Sebastiani 2006). In this phase, attitude idea and feelings of people in each tweet are specified. For this purpose, the open-source software PHPInsight is used which is designed to determine negative or positive attitudes of tweets is used (Github 2016). PHPInsight is involved in two processes of filtering and classifying, while filtering it makes sure that only words are used in processing, it then does the classifying. The software uses Sentiwordnet 3.0 (Baccianella et al. 2010) dictionary for putting the tweets in to three groups of positive negative and neutral. It then does the sorting, using the Bayes algorithm. For the end, first we dropped mentions, hashtags, URLs, and stop words from our dataset using custom queries in MySQL. Then the pruned dataset fed into PHPinsight. The software associated the input tweet t to three numerical scores Pos(t), Neg(t), Obj(t) which indicate how positive, negative, and "objective" (i.e., neutral). The maximum score for each tweet represents the feeling of the tweet and the tweet labeled accordingly.

3.6 Analyzing geographical location

Users' geographical location can be studies using the fields below:

1. Geographical latitude and longitude (geo_lat, geo_long) fields: If the users have sent their tweets using devices capable of GPS accessory, this fields are utilizable.
2. Location field: The most common way present in users' profile is location. This field is a free-text string entered by the user.
3. Timezone field: While the location field is a free-text string entered by the user, the time zone field is a selection made from a drop-down menu. The time zone entries consist of a location name alongside a UTC offset, which can be used to determine the user's country of residence.

Table 3 Key words for sorting out topics

Diet	Dailydiet, lifestyle, naturalnfit, loseweight, healthyliving, fast solution, nutrition, weightloss, fatloss, tips, program, dailyprogram, mealplanning, planning, Diet, Dietplan, dailydiet, tip, planning
Gym	Sport, gymlife, gym, exercise, walk, bike ride, gymlife, gymtime, sport, ride, workout, lifestyle, training, bodybuilding, running, muscle, fitnessmodel, crossfit, sport, run, fitnesslife, health, yoga, benefits, fitfam, fitbit, wellness
Disease	Disease, epidemic, disease, illness, cancer, die, dies, died, diabete, sickness, diabetes, people, disease, risk, cancer, holistic, epidemic
Fundraising	Fundraising, money, fundraiser, fundraising, donate, donating, £, $, fund, donation, proceeds, benefit, pay, expense, charity, charities, raise.*research, support.*research, raise
Motivation	Motivation, beautiful, fitness, fashion, fit, workout, motivation, body life, sexy, hot, fitnessaddict, babe, fitspo, love, fitfam, goals, butt, fitnessmotivation, sex
Obesity	Obesity, comic, cartoon, disney, smiley, smiles, photographer, superhero cape, lily pad, artist, song

The stages of this analysis are as follows; first the users' location is investigated. For this end, timezone field has priority. In case of containing a value, the location of user is determined using timezone field. For the Twitter users with an empty timezone field, the location field is considered for determining user's location. The reason for giving priority to timezone field is that location field is prone to contain misleading information which is filled by Twitter users who do not want to reveal their real location.

In the next step, cities and countries are classified in to five continents of America, Europe, Asia, Africa, and Oceania. And the geographical location of the tweets sent by users was identified. If none of the tweets sent by one user have got a country name, the word unknown was labeled for the geographical location of the user.

4 Result

The number of received tweets from the date December 27,2017 to January 30, 2018 was 2,684,858 which was available due to the 5 weeks. To analyze the data, the software SPSS was used. In this paper, considering the difference between various groups in the graphs, we used statistical tests to examine the significance of these differences. Given that some of the variables were at the ordinal level, Chi-square test is used. We have used ANOVA test for interval variables due to the analysis of variance assumptions. The results and analysis of the data obtained in the previous phase are presented here.

4.1 Results of the analysis of the time zone

After the pre-process phase, 1,673,559 tweets remained in dataset. To determine the days with the largest number of tweets, a daily comparison of the number of sent tweets was made during the time period December 27, 2017 to January 30, 2018. Figure 1 shows the number of sent tweets during 5 weeks divided by week days. Figure 2 shows the total number of tweets for every week day.

As Fig. 2 indicates the frequency of sent tweets on Tuesdays to Fridays is higher compared to other days of the week, with Saturday having the least number.

The monthly average of the number of sent tweets related to weight loss and fitness is 47,815 tweets. Chi-square test suggested a significant difference between the number of sent tweets and the days of week ($\chi^2 = 2686.68$, $p < 0.005$).

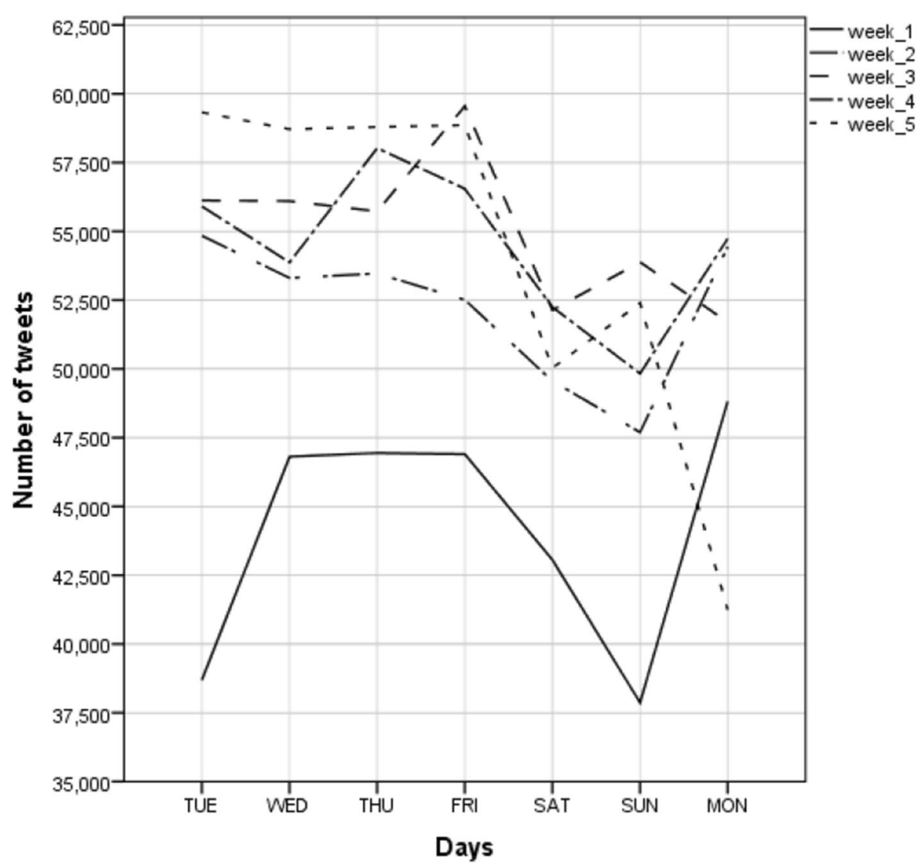

Fig. 1 Frequency of tweets in different weeks divided by weekdays

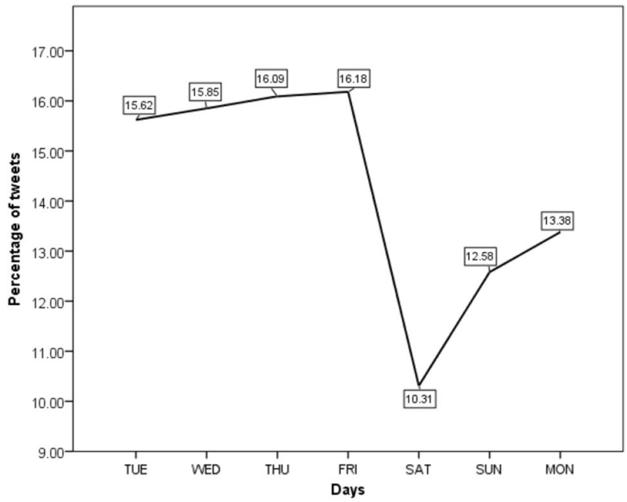

Fig. 2 Aggregate frequency of tweets divided by weekdays

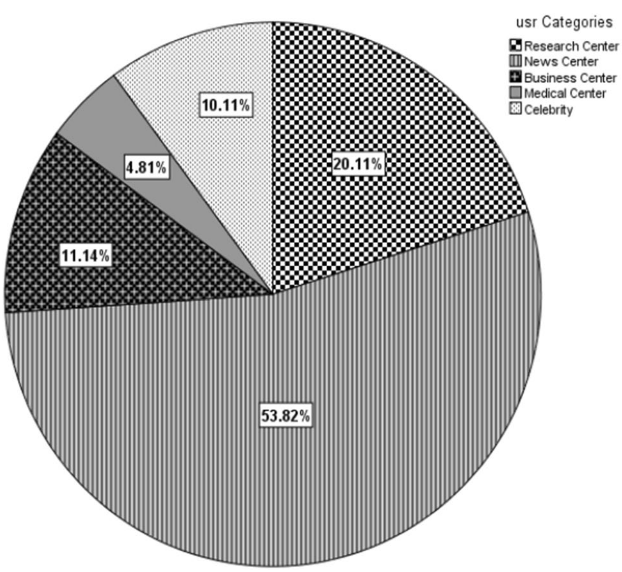

Fig. 3 Percentage of users other than common people in five groups

4.2 Results of the analyzing users

As mentioned previously, in the present study, users sending tweets related to weight loss and fitness are divided into six groups, which made a population of 355,856 users.

97.9% of users belong to the common people group and the rest 2.1% of users, which are a population of 7584 accounts are divided in five groups of research center, news center, business center, medical center, and celebrity, which are depicted in Fig. 3.

As Fig. 3 indicates, news and research centers are more active in the field of weight loss. Table 4 indicates the number of original tweets and retweets for each user group. Table 4 indicates that during the period of this study, the tweets for research center were original ones and news center mostly had retweets. Chi-square test suggested that there is a significant difference between the number of tweets and retweets and the grouping of users ($\chi^2 = 1533.876$, $df = 5$, $p < 0.005$).

Figure 4 shows an analysis of weekdays and the activities of user groups with common people excluded. As can be seen from the figure, news centers have the highest level of activity and medical centers have the lowest activity. Although medical centers are expected to be active in the field, the reason for this difference may be due to the different way medical centers are advertised. This difference suggests that these centers are doing something different in the field of weight loss, such as web sites and Instagram, which can be explored in future work. Chi-square test suggested that there is a significant difference between weekdays and users' activity level ($\chi^2 = 422.755$, $p < 0.005$).

Figure 5 demonstrates the frequency of tweets based on users' geographical location and user groups. As shown in the Fig. 5, the most proportion of tweets related to weight loss and fitness belongs to America. The subsequent places are Europe, Asia, Oceania, and Africa. The amount of activity similarly in all continents is for news centers with the highest followed by research centers, business centers, celebrities and medical centers. Chi-square test sets out a significant difference between the activities of users' groups and their geographical location ($\chi^2 = 7069.229$, $p < 0.005$).

Table 4 Comparing the number of tweets and users' groups

	User categories						Total
	Individual	Research center	News center	Business center	Medical center	Celebrity	
Tweet	1,190,150	12,152	28,068	7989	1655	3982	1,243,996
	74.3%	79.9%	75.8%	80.4%	85.2%	58.2%	74.3%
Retweet	412,440	3058	8976	1944	288	2857	429,563
	25.7%	20.1%	24.2%	19.6%	14.8%	41.8%	25.7%
Total	1,602,590	15,210	37,044	9933	1943	6839	1,673,559
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

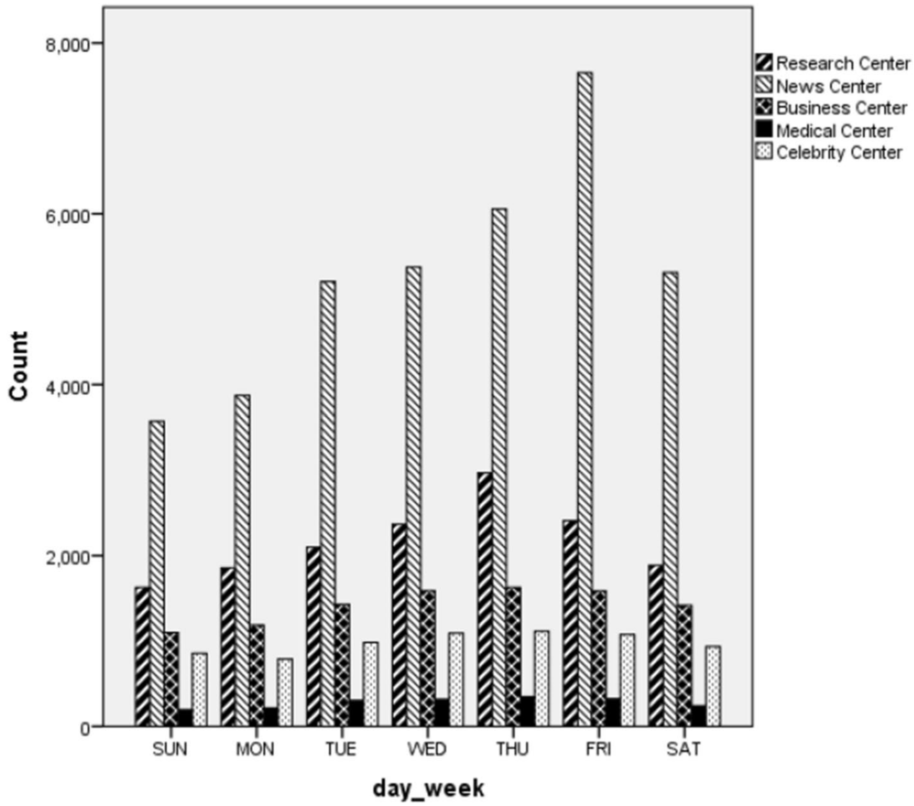

Fig. 4 Activity level of different groups in weekdays

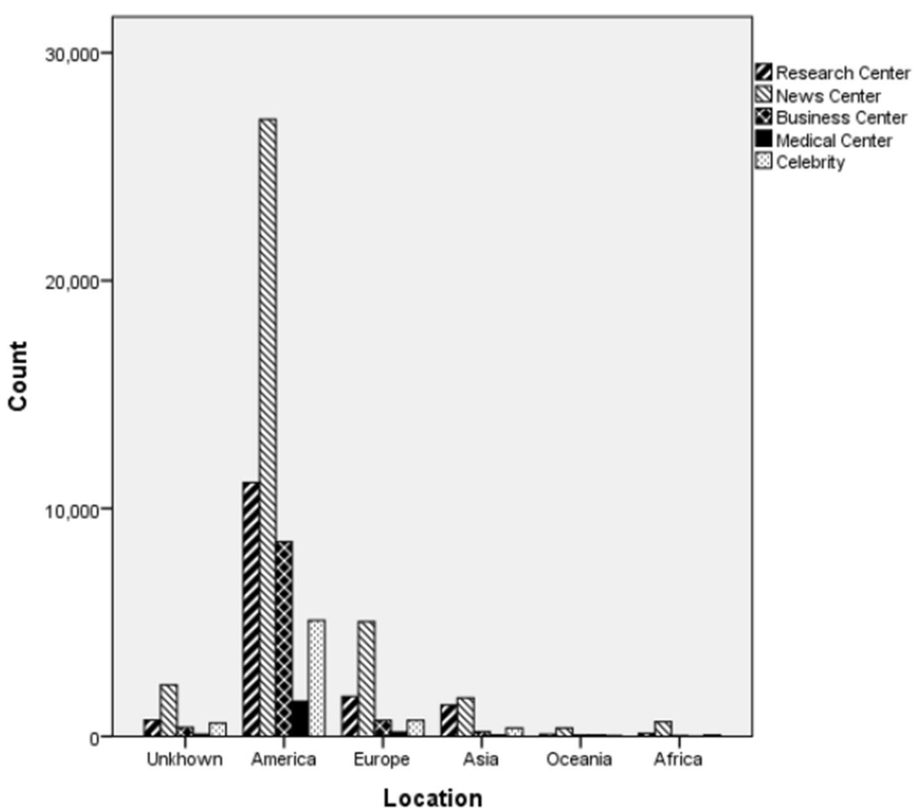

Fig. 5 Frequency of tweets divided by geographical location and user groups

Users in the common user' group sent 1,602,590 tweets, of which 63.8% belongs to America, 14.8% to Europe, 4.7% to Asia, 1.1% to Oceania and 1.2% to Africa. 14.4% of tweets location is unspecified.

Table 5 provides the statistical data through an ANOVA model. ANOVA model suggested that is a significant relation between users' groups and the number of followers ($\chi^2 = 18{,}588.483$, $p < 0.005$).

4.3 Results of tweets analysis

MALLET software is employed to categorize tweets' topics, which benefits the LDA algorithm. Figure 6 illustrates the frequency of tweets in every topic.

As shown in Fig. 6, the motivation topic appears most in the tweets, and the topic with the least number of tweets is investment. 0.1% of tweets could not be categorized. The next step was to analyze the feeling of tweets in every topic. From all the topics, 57.1% are neutral, 30.7% positive, and 17.1% negative. Figure 7 shows positive and negative feeling output analysis of each topic.

As can be seen from Fig. 7, motivation surpass the topics in terms of the number of positive and negative tweets.

To study the number of followers in each topic and to find out whether positive or negative topics have been followed more, the average of the number of followers in each topic is calculated and shown in Fig. 8.

Figure 8 indicates that the number of followers in topics with negative feelings related to obesity is more than others; also the number of followers for positive topics is highest although investment topics have the least number of tweets with regards to number. Analysis of two factor variance of the two independent variables of topic and feeling for each tweet is significant on the number of followers ($\chi^2 = 14.271$, $p < 0.005$).

Chi-square test set out a significant difference between the feeling analysis of each tweet and topics ($\chi^2 = 50{,}345.582$, $p < 0.005$).

Table 6 illustrates the number of tweets in each topic along with its geographical location. It also shows each continent's share in the creation of each topic. To have a more precise analysis of the identified topics, the number of tweets is compared in terms of feelings viewpoint in daily time periods. Figure 9 depicts the graphs of each topic with attitudes on different weekdays. As can be seen from the figure, frequency of positive tweets with diet content, is generally slightly more than negative tweets, whereas the difference between number of positive and negative tweets with gym suggestion content is considerable. On Wednesdays and Fridays, the number of topics with a positive attitudes related to disease is almost the same as the negative ones, whereas negative attitudes dominate on Tuesdays and on other days the dominant attitudes are positive. Since the difference between negative and positive attitudes is negligible in most days of week, no explanation can be found for the dominancy of positive or negative attitudes.

Table 5 ANOVA model for the dependent variable of number of followers and users' groups

	Sum of squares	*df*	Mean square	*F*	Sig.
Between groups	5.083E+14	5	1.017E+14	18,588.483	0.000
Within groups	9.152E+15	1,673,553	5,468,806,628		
Total	9.661E+15	1,673,558			

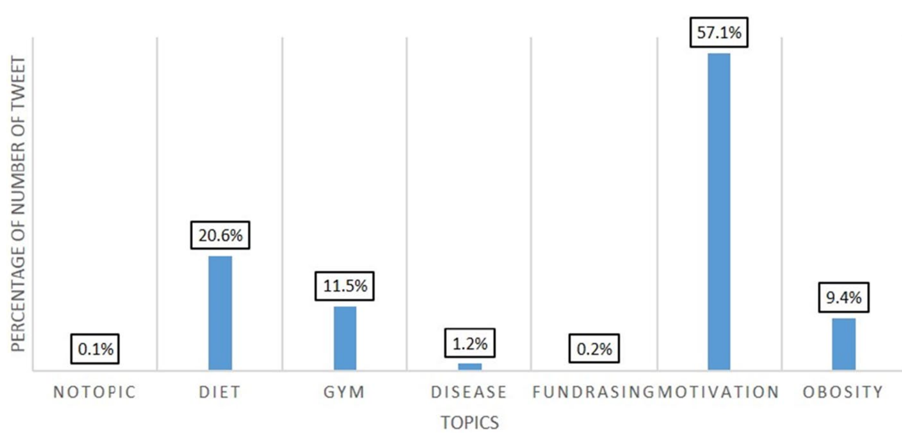

Fig. 6 The frequency of tweets in each topic

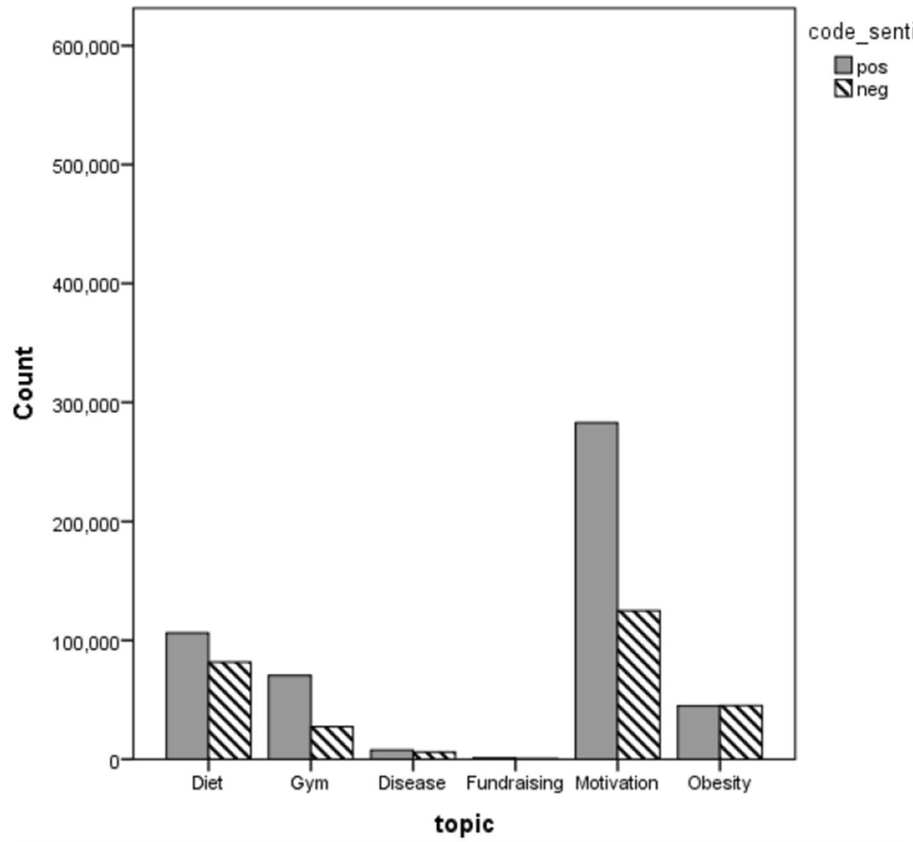

Fig. 7 Frequency of positive and negative tweets in each topic

5 Discussion

This study empirically examined a large data from Twitter; it is found that the continents with the most contribution in this respect are America Europe and Asia. There are not many organizations active in the field in different countries. Compared to the high activity of American and European users, the number of these organizations is handful. It is possible to increase people awareness and make good suggestions for people's health by establishing organizations active in the field of weight loss and fitness.

The most quoted discussion in the field of weight loss and fitness is motivation which means commercial and pharmaceutical centers can make use of this point to motivate their users.

With regards to weight loss illnesses, we analyzed the topic related to illnesses using LDA method and extracted the recurrent words from them in this topic, the illnesses mostly referred diabetes high blood pressure and heart attack. Weight loss was offered as a way to prevent such diseases.

This research was carried out in a period of 1 month and it is possible to expand this time period for a more detailed analysis and extract more of the mentioned samples.

A more accurate analysis on user's geographical location can be reached by increasing time periods and considering the aforementioned facts.

6 Conclusion

In the present study, we have analyzed the data related to weight loss from December 27, 2016 and January 30, 2018. After studying the data and results, it was found that there is a significant relation between topic, geographical location, positive–negative attitudes, and the number of followers. It was observed that the problem of obesity is still an issue of concern to many; there are high motivations for overcoming this problem. We also found that not many organizations are active in this field and that by increasing the number of such organizations with effective daily tweets, we can expect an improvement in weight loss. For improving this research, we can expand its time period. There could also be an analysis of the age of people dealing with the problem of weight loss. The topic of motivation for slimness could also be categorized and, therefore, form a clearer understanding of people's motivations in this respect.

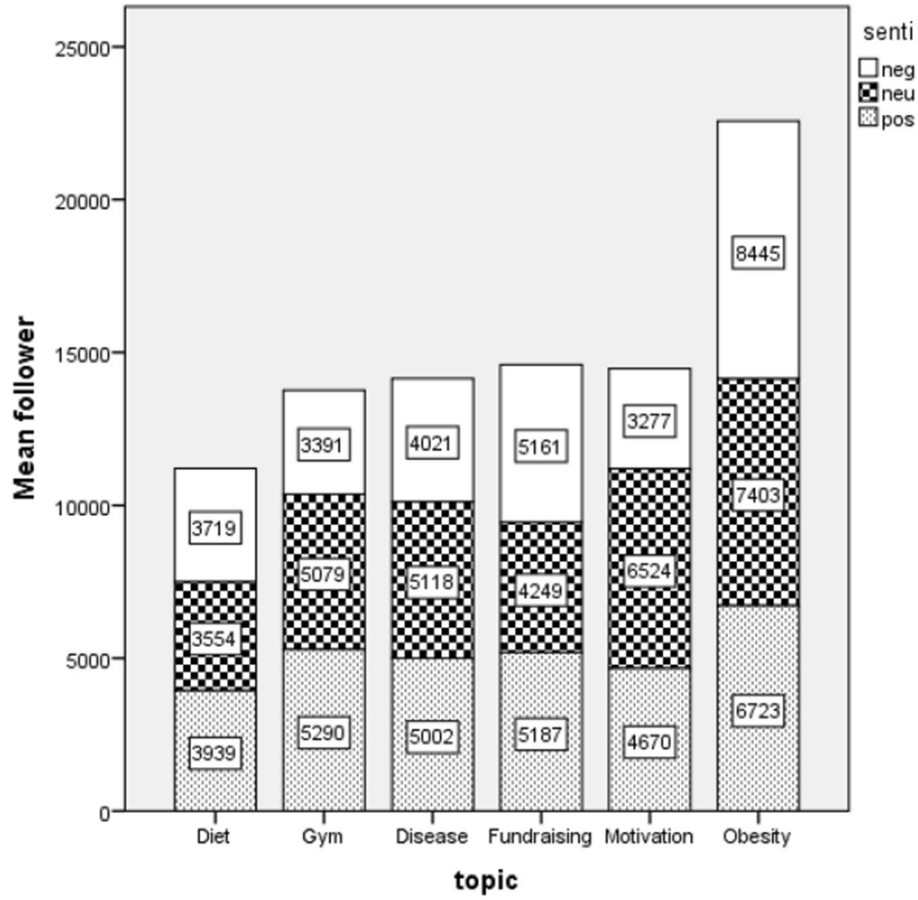

Fig. 8 Number of followers in each topic

Table 6 Frequency of tweets in each continent divided by topics

	Unknown	America	Europe	Asia	Oceania	Africa	Total
Notopic	288	1272	304	83	46	23	2016
	14.3%	63.1%	15.1%	4.1%	2.3%	1.1%	100.0%
Diet	52,114	223,861	44,960	17,220	3426	3388	344,969
	15.1%	64.9%	13.0%	5.0%	1.0%	1.0%	100.0%
Gym	26,723	123,800	27,300	9103	2571	2510	192,007
	13.9%	64.5%	14.2%	4.7%	1.3%	1.3%	100.0%
Disease	2108	12,899	2825	1057	303	279	19,471
	10.8%	66.2%	14.5%	5.4%	1.6%	1.4%	100.0%
Fundraising	355	1832	479	119	42	24	2851
	12.5%	64.3%	16.8%	4.2%	1.5%	8.%	100.0%
Motivation	133,212	614,693	142,996	42,595	9821	12,076	955,393
	13.9%	64.3%	15.0%	4.5%	1.0%	1.3%	100.0%
Obesity	19,458	97,334	26,160	8571	2733	2596	156,852
	12.4%	62.1%	16.7%	5.5%	1.7%	1.7%	100.0%
Total	234,258	1,075,691	245,024	78,748	18,942	20,896	1,673,559
	14.0%	64.3%	14.6%	4.7%	1.1%	1.2%	100.0%

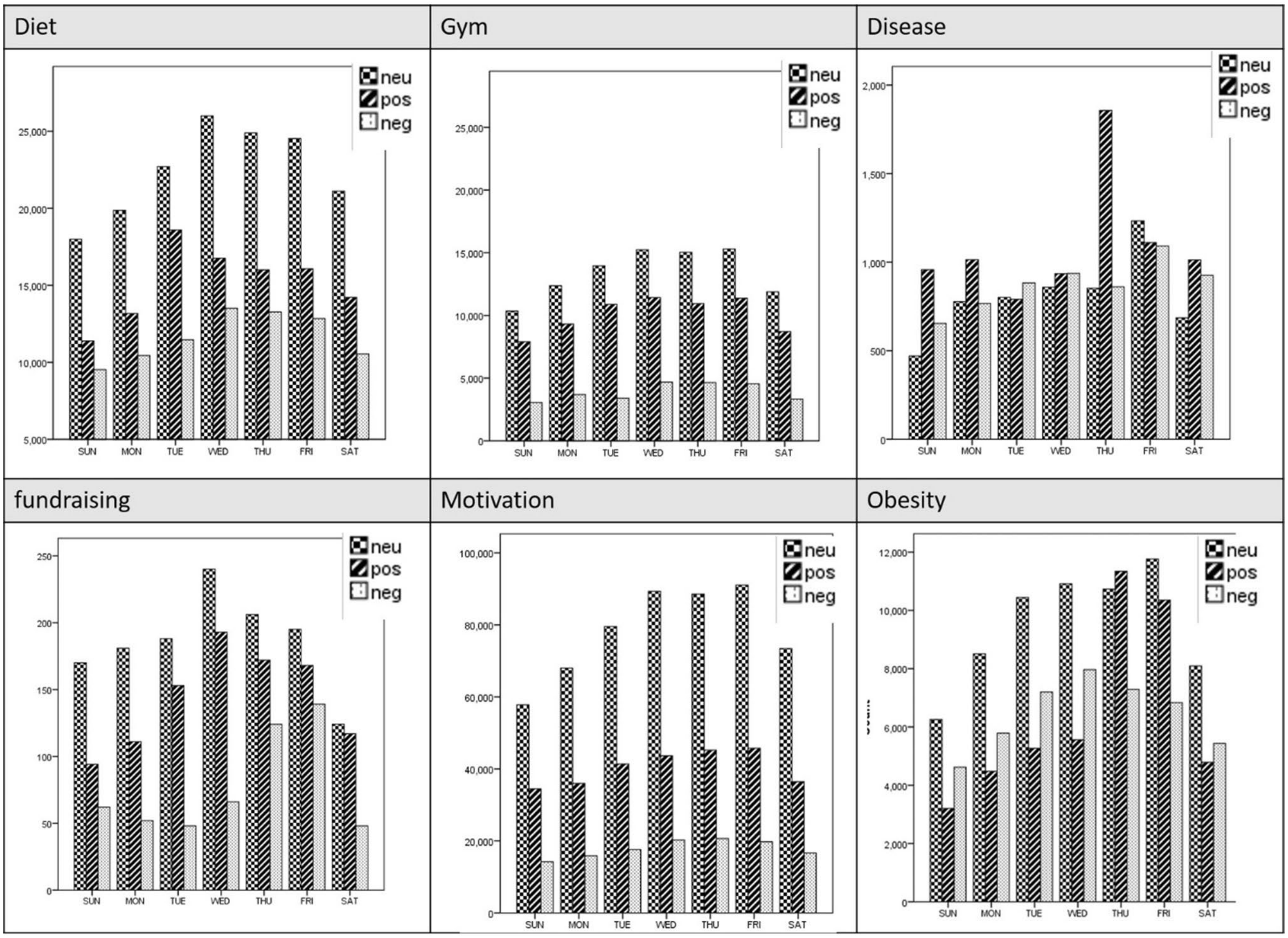

Fig. 9 Comparison of attitudes of tweets divided by topics and weekdays

Aknowledgement This paper derives from the Research Project with code 98-1-37-14862 and Approval ID IR.IUMS.REC.1398.308.

Author contributions MY studied conception and design, data collection, drafting of manuscript, drafting of manuscript, analysis and interpretation of data, and writing the article. MH participated in study conception and design, analysis and interpretation of data, final approval of article. MD participated in analysis and interpretation of data, writing the article, and final approval of article.

Funding Not applicable.

Availability of data and materials The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflict of interests The authors declare that they have no competing interests.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

References

Brandwatch (2016) https://www.brandwatch.com/blog/96-amazing-social-media-statistics-and-facts-for-2016/. Accessed 27 Dec 2016

Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the seventh international conference on language resources and evaluation (LREC'10). European Language Resources Association (ELRA), pp 2200–2204

Blei DM (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022

Chou SW-Y (2009) Social media use in the United States: implications for health communication. J Med Internet Res 11(4):e48. https://doi.org/10.2196/jmir.1249

Esuli A, Sebastiani F (2006) Sentiwordnet: a publicly available lexical resource for opinion mining. In: The international conference on language resources and evaluation, no 6, pp 417–422

Gabrielle M, Turner-McGrievy M (2015) Tweet for health: using an online social network to examine temporal trends in weight loss-related posts. Transl Behav Med 5(2):160–166
Github (2016) https://github.com/JWHennessey/phpInsight. Accessed 12 Dec 2016
Gruver SR-G (2016) A social media peer group intervention for mothers to prevent obesity and promote healthy growth from infancy: development and pilot trial. JMIR Res Protoc 5(3):e159. https://doi.org/10.2196/resprot.5276
Lydecker JC (2016) Does this tweet make me look fat? A content analysis of weight stigma on twitter. Eat Weight Disord 21(2):229–235
Martinus Evans P (2016) The weight loss blogosphere: an online survey of weight loss bloggers. Transl Behav Med 6(3):403–409
May CN, Waring ME, Rodrigues S, Oleski JL, Olendzki E, Evans M, Carey J, Pagoto SL (2016) Weight loss support seeking on twitter: the impact of weight on follow back rates and interactions. Transl Behav Med 7(1):84–91
McCallum AK (2002) MALLET: a machine learning for language toolkit. http://mallet.cs.umass.edu. Accessed 12 Dec 2016
Richard Dobbs CS (2014) Overcoming obesity: An initial economic analysis. McKinsey Global Institute, New York
Safran Naimark JM (2015) The impact of a web-based app (eBalance) in promoting healthy lifestyles: randomized controlled trial. J Med Internet Res 17(3):e56
Waring ME (2016) Interest in a twitter-delivered weight loss program among women of childbearing age. Transl Behav Med 6(2):277–284